Bayesian nonparametric modeling approaches for quantile regression

Athanasios Kottas

Department of Applied Mathematics and Statistics

University of California, Santa Cruz

Department of Statistics

Athens University of Economics and Business

December 19, 2007

Outline

1. Introduction and motivation
2. Dirichlet process priors and Dirichlet process mixtures
3. Bayesian semiparametric quantile regression
4. Fully nonparametric inference for quantile regression

1. Introduction and motivation

- In regression settings, the covariates may have effect not only on the center of the response distribution but also on its shape
- Quantile regression quantifies relationship between a set of quantiles of response distribution and covariates, and thus, provides a more complete explanation of the response distribution in terms of available covariates
- Applications: econometrics, medicine, social sciences, educational studies ...
- Objective is to develop modeling for quantile regression that:
  → relaxes, as much as possible, parametric assumptions
  → enables full and exact inference for the quantile regression function and any other feature of the response distribution that might be of interest

Introduction and motivation

- The area of **Bayesian nonparametrics** provides the framework for such modeling
  → instead of specifying unknown functions and distributions up to a (small) number of parameters, treat them as the random model parameters
  → *nonparametric priors* support the underlying spaces of random functions/distributions resulting in flexible inferences given the data

- Two modeling approaches to quantile regression:
  → a semiparametric model where the error distribution is assigned a nonparametric prior and the regression function is modeled parametrically
  → a fully nonparametric approach where the joint distribution of the response and the covariates is modeled with a mixture model, with posterior inference for quantile curves emerging through the conditional distribution of the response given the covariates

- Both approaches utilize Dirichlet process mixtures, a flexible class of nonparametric mixture models

2. Dirichlet process priors and Dirichlet process mixtures

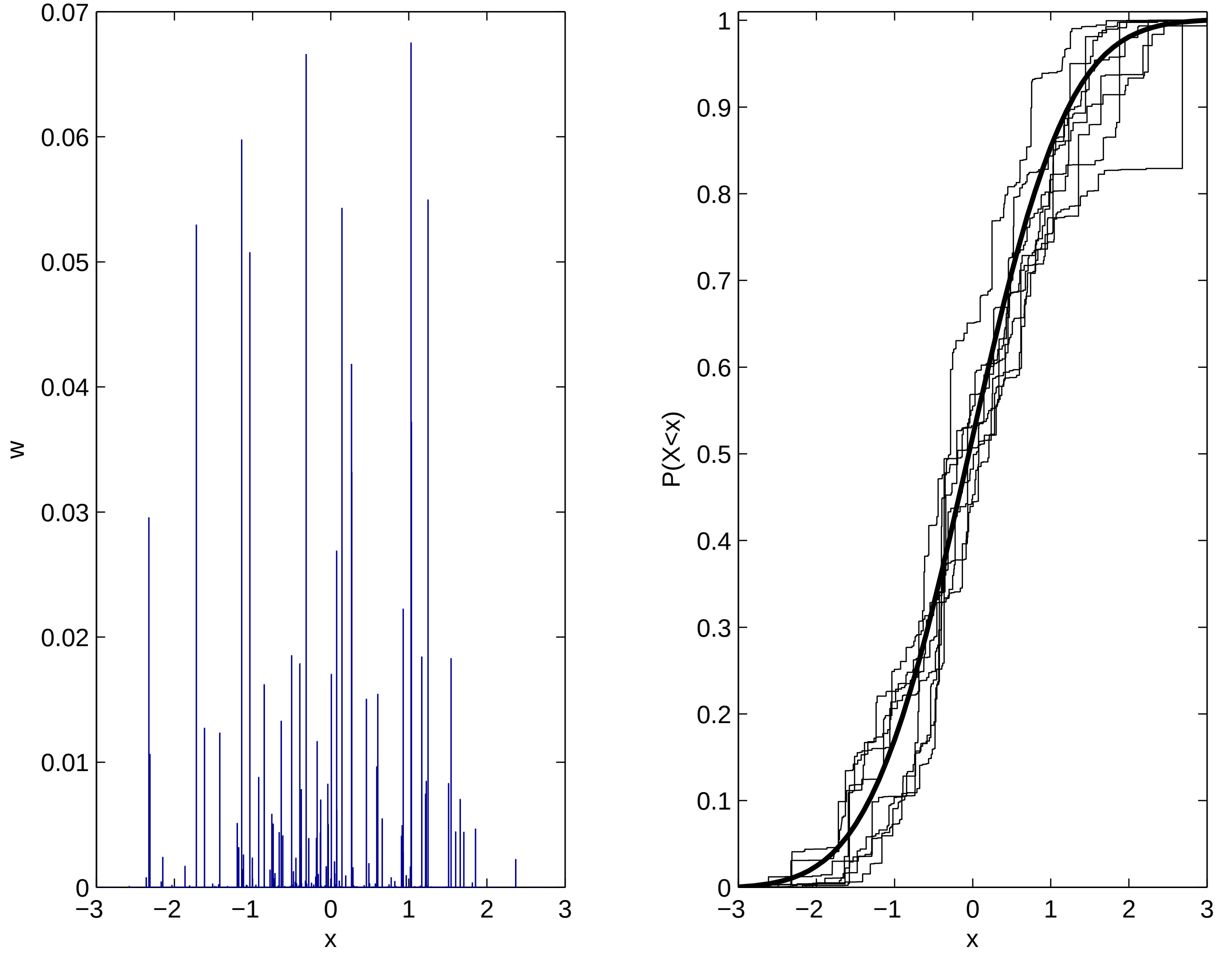
- The Dirichlet process (DP) (Ferguson, 1973) is a random probability measure on distributions characterized by two parameters: a base distribution G_0 (the center of the process) and a (precision) parameter $\alpha > 0$

- DP constructive definition (Sethuraman, 1994)
  → let $\{z_s, s = 1, 2, ...\}$ and $\{\phi_j, j = 1, 2, ...\}$ be independent sequences of random variables, with z_s i.i.d. Beta$(1, \alpha)$, and ϕ_j i.i.d. G_0
  → define $\omega_1 = z_1$, $\omega_j = z_j \prod_{s=1}^{j-1}(1 - z_s)$, $j \geq 2$ (*stick-breaking* construction)
  → then, a realization G from DP(α, G_0) is (almost surely) of the form

$$G(\cdot) = \sum_{j=1}^{\infty} \omega_j \delta_{\phi_j}(\cdot)$$

  i.e., a discrete distribution that can be represented as a countable mixture of point masses

Dirichlet process priors and Dirichlet process mixtures



DP with $G_0 = N(0,1)$ and $\alpha = 20$. In the left panel, the spiked lines are located at 1000 sampled values of x drawn from $N(0,1)$ with heights given by the weights, ω_ℓ, calculated using the stick-breaking algorithm (a truncated version so that the weights sum to 1). These spikes are then summed from left to right to generate one cdf sample path from the DP. The right panel shows 8 such sample paths indicated by the lighter jagged lines. The heavy smooth line indicates the $N(0,1)$ cdf.

Dirichlet process priors and Dirichlet process mixtures

- **Dirichlet process mixture model:** for a parametric family of distributions $K(\cdot;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$, define

$$F(\cdot;G) = \int K(\cdot;\boldsymbol{\theta})\mathrm{d}G(\boldsymbol{\theta}), \quad G \sim \mathrm{DP}(\alpha, G_0)$$

$\rightarrow$ DP mixture prior can model both discrete and continuous distributions (e.g., $K(\cdot;\theta)$ might be Poisson, binomial, normal, gamma, multivariate normal, ...)

- **Hierarchical model:** for $y_1, ..., y_n$ i.i.d., given G, from $F(\cdot;G)$,

$$\begin{aligned} y_i \mid \boldsymbol{\theta}_i &\overset{ind.}{\sim} K(\cdot;\boldsymbol{\theta}_i),\ \ i = 1, ..., n \\ \boldsymbol{\theta}_i \mid G &\overset{i.i.d.}{\sim} G,\ \ i = 1, ..., n \\ G &\sim \mathrm{DP}(\alpha, G_0) \end{aligned}$$

$\rightarrow$ typically, hyperpriors on α and/or the parameters ψ of $G_0 \equiv G_0(\psi)$ are added

- Posterior simulation methods (mainly MCMC) for $p(\boldsymbol{\theta}, \alpha, \psi, G \mid \text{data})$

3. Bayesian semiparametric quantile regression

(joint work with Milovan Krnjajić)

- Response observations y_i, with covariate vectors $\boldsymbol{x}_i$, $i = 1, ..., n$
- Additive quantile regression formulation: $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$
  $\rightarrow$ ε_i i.i.d. from an error distribution with p-th quantile equal to 0

- **Parametric modeling:** specifies parametrically the error distribution
  $\rightarrow$ e.g., asymmetric Laplace distribution (**model** $\mathcal{M}_0$):

$$\varepsilon_i \stackrel{iid}{\sim} k_p^{AL}(\varepsilon; \sigma) = \sigma^{-1} p(1-p) \exp\{-\sigma^{-1}\varepsilon(p - 1_{(\sigma^{-1}\varepsilon < 0)})\}$$

  with $\int_{-\infty}^{0} k_p^{AL}(\varepsilon; \sigma) \mathrm{d}\varepsilon = p$

- Limitation: one parameter p determines both quantile and skewness ($p > 0.5$ left skewed, $p = 0.5$ symmetric, $p < 0.5$ right skewed) — for example, the error distribution is symmetric in the median regression case

- **Objective**: develop flexible nonparametric prior models for the random error density $f_p(\cdot)$

DP mixture models for the quantile regression error density

- **Model** $\mathcal{M}_1$: general scale mixture of asymmetric Laplace densities

$$f_p^1(\varepsilon; G) = \int k_p^{AL}(\varepsilon; \sigma) \mathrm{d}G(\sigma), \qquad G \sim \mathrm{DP}(\alpha, G_0)$$

$\rightarrow$ captures more flexible tail behavior (mixing preserves quantiles, $\int_{-\infty}^{0} f_p^1(\varepsilon; G) \mathrm{d}\varepsilon = p$)

- $\mathcal{M}_1$ extends $\mathcal{M}_0$ with regard to tail behavior, but the skewness of the mixture $f_p^1(\cdot; G)$ suffers the same limitation as the kernel $k_p^{AL}(\cdot; \sigma)$

- A key result that allows a more flexible model formulation is a *representation* theorem for non-increasing densities on R^+:
$\rightarrow$ For any non-increasing density $f(\cdot)$ on R^+ there exists a distribution function G, with support on R^+, such that $f(t;G) = \int \theta^{-1} 1_{[0,\theta)}(t) \mathrm{d}G(\theta)$

- This result leads to a mixture representation for *any* unimodal density on the real line with p-th quantile (and mode) equal to zero,

$$\iint k_p(\varepsilon;\sigma_1,\sigma_2)\mathrm{d}G_1(\sigma_1)\mathrm{d}G_2(\sigma_2)$$

with G_1 and G_2 supported by R^+, and

$$k_p(\varepsilon;\sigma_1,\sigma_2) = \frac{p}{\sigma_1} 1_{(-\sigma_1,0)}(\varepsilon) + \frac{(1-p)}{\sigma_2} 1_{[0,\sigma_2)}(\varepsilon),$$

with $0 < p < 1$, $\sigma_r > 0$, $r = 1, 2$

- Assuming independent DP priors for G_1 and G_2, we obtain **model** $\mathcal{M}_2$:

$$f_p^2(\varepsilon; G_1, G_2) = \iint k_p(\varepsilon; \sigma_1, \sigma_2) \mathrm{d}G_1(\sigma_1) \mathrm{d}G_2(\sigma_2), \; G_r \sim \mathrm{DP}(\alpha_r, G_{r0}), r = 1, 2$$

$\rightarrow$ model $\mathcal{M}_2$ can capture general forms of skewness and tail behavior

- The full hierarchical model $\mathcal{M}_2$:

$$\begin{aligned} y_i \mid \boldsymbol{\beta}, \sigma_{1i}, \sigma_{2i} &\overset{ind}{\sim} k_p(y_i - \boldsymbol{x}_i' \boldsymbol{\beta}; \sigma_{1i}, \sigma_{2i}), \;\; i = 1, ..., n \\ \sigma_{ri} \mid G_r &\overset{iid}{\sim} G_r, \; r = 1, 2, \;\; i = 1, ..., n \\ G_r \mid \alpha_r, d_r &\sim \mathrm{DP}(\alpha_r, G_{r0} = \mathrm{IGamma}(c_r, d_r)), \; r = 1, 2 \end{aligned}$$

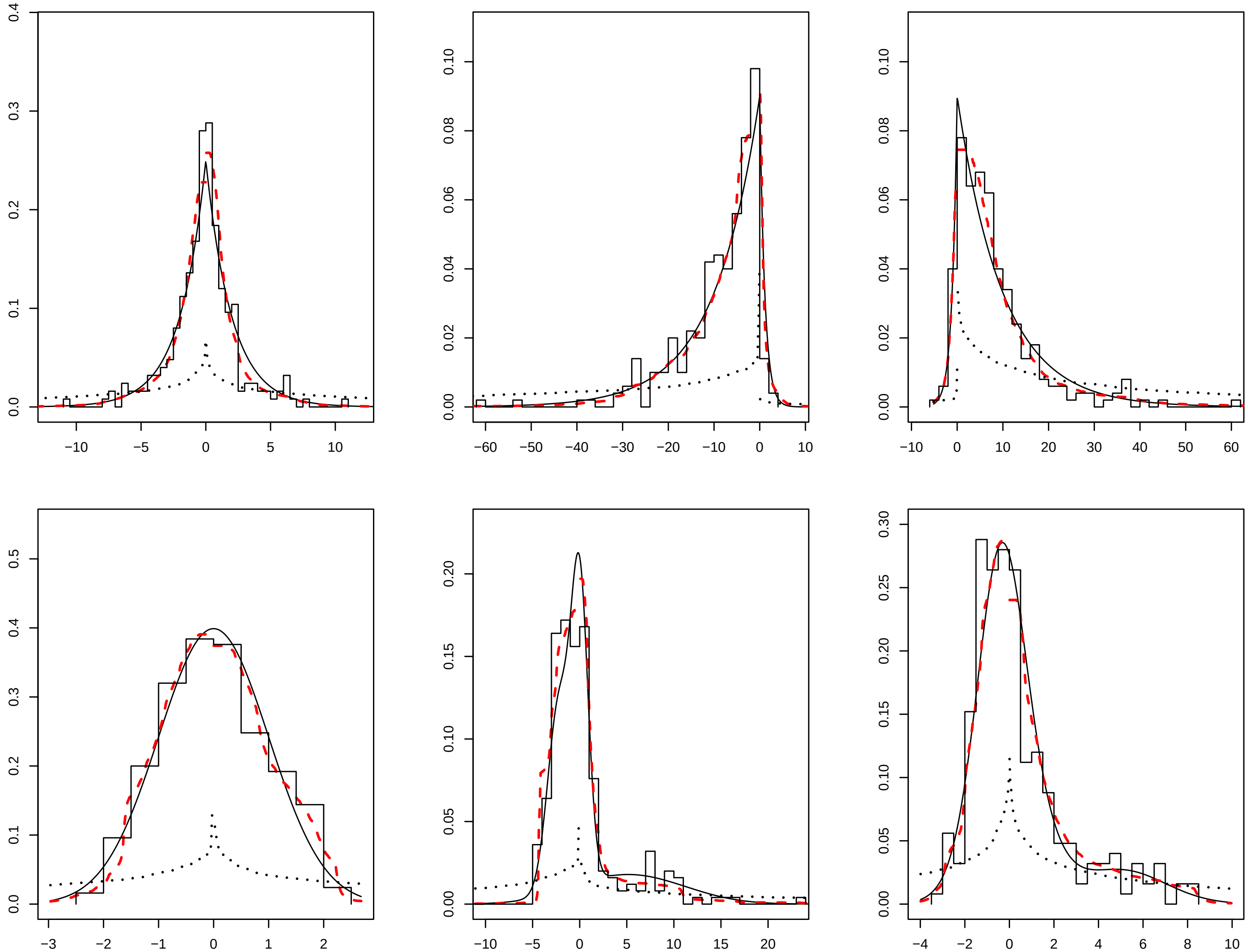
- Posterior inference under all models is obtained using standard MCMC methods for DP mixture models (censoring can also be handled)

Data Illustrations

- Simulated data ($n = 250$ in each case) from distributions with a specific quantile fixed at 0 (no covariates) and with varying shapes
  $\rightarrow$ three standard Laplace distributions ($\sigma = 1$) for three values of p ($p = 0.5$, 0.9, and 0.1)
  $\rightarrow$ a standard normal distribution, and two mixtures of normals, one with 0.6-th quantile at zero and another with median zero (the components for both normal mixtures are chosen so that the resulting mixture densities are right skewed with non-standard tail behavior)

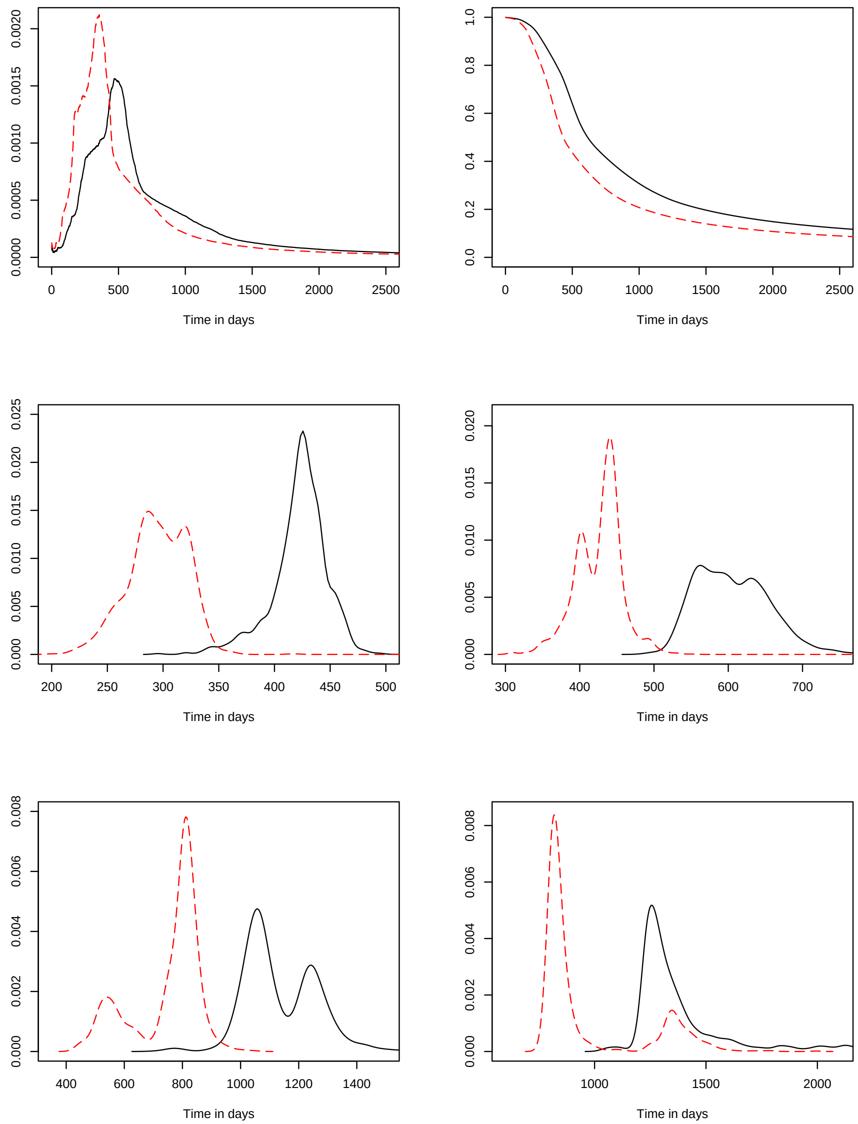
- Small cell lung cancer data: survival times in days for 121 patients with small cell lung cancer; 23 survival times are right censored
  $\rightarrow$ each patient was randomly assigned to one of two treatments A and B, achieving 62 and 59 patients, respectively (treatment indicator is the covariate)

Bayesian semiparametric quantile regression



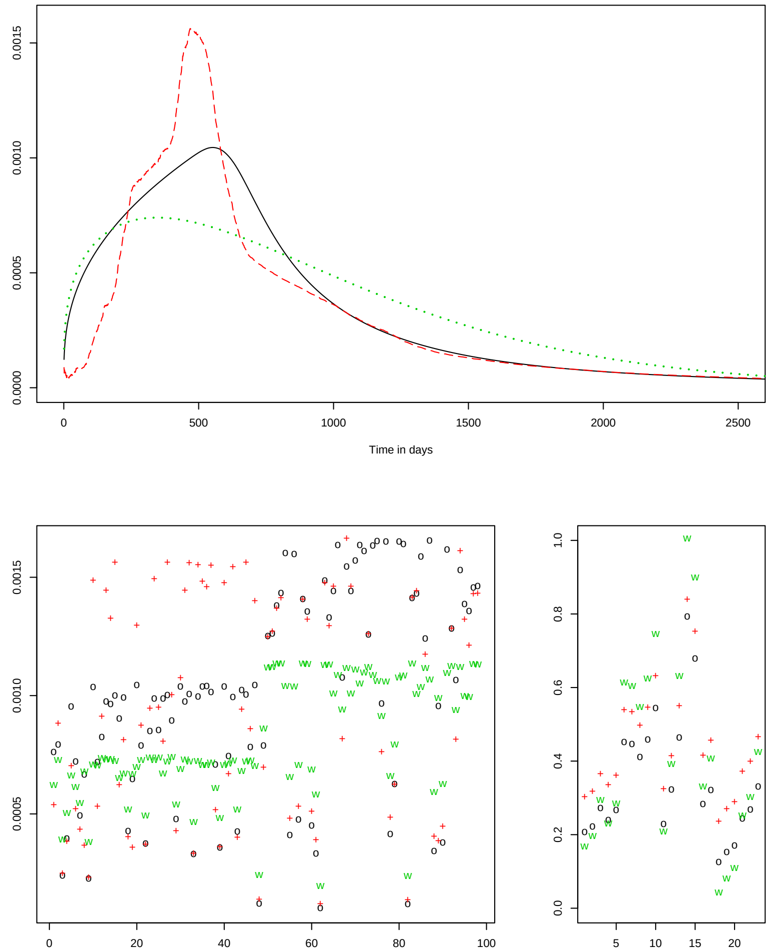
Simulation study. Prior and posterior predictive densities (dotted and dashed lines) under model $\mathcal{M}_2$. The solid lines denote the true densities; the histograms of the data are also included.

Bayesian semiparametric quantile regression



Small cell lung cancer data. Model $\mathcal{M}_2$ posterior predictive densities and survival functions, and posteriors for 25th, 50th, 75th and 90th percentile survival times for treatment A and B (solid and dashed lines).

Bayesian semiparametric quantile regression



Small cell lung cancer data. The top panel displays posterior predictive densities for treatment A under model $\mathcal{M}_0$ (solid line), model $\mathcal{M}_2$ (dashed line), and a parametric Weibull model (dotted line). The bottom panels include CPO plots for the uncensored (left panel) and censored data (right panel). The "o" denote CPO values under model $\mathcal{M}_0$, "+" under model $\mathcal{M}_2$, and "w" under the Weibull model.

Quantile regression with dependent error densities

- **Motivation**: Under the previous setting, the distribution of ε_i is the same for all x_i, and thus, the distribution of y_i changes with x_i only through the p-th quantile $\boldsymbol{x}_i'\boldsymbol{\beta}$

- To model nonparametrically error distributions that change with covariates, we need a prior model for

$$f_{p,\mathcal{X}}(\cdot) = \{f_{p,x}(\cdot) : x \in \mathcal{X}\},$$

where $\mathcal{X}$ is the covariate space, and for each x, $\int_{-\infty}^{0} f_{p,x}(\varepsilon)\mathrm{d}\varepsilon = p$

- For example, under model $\mathcal{M}_2$, to allow $f_p^2(\varepsilon; G_1, G_2)$ to change with x, the mixing distributions G_1, G_2 need to change with x — for $r = 1, 2$, we need to replace G_r with a stochastic process $G_{r,\mathcal{X}}$ over $\mathcal{X}$

Bayesian semiparametric quantile regression

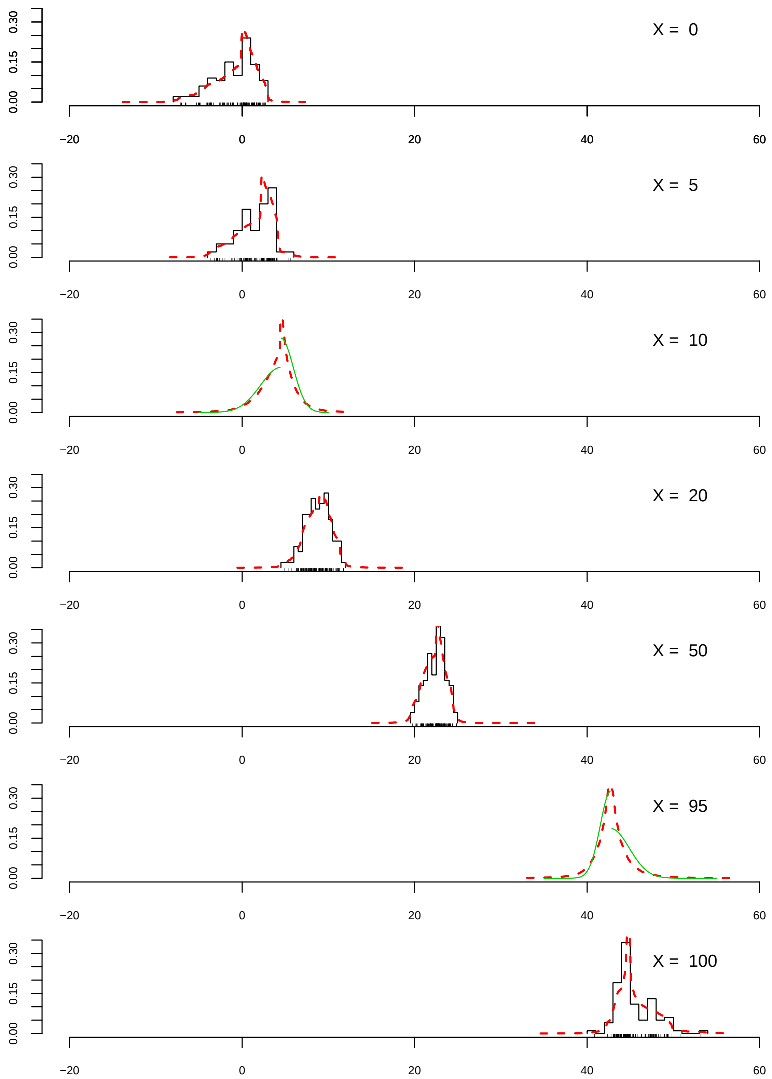
- Dependent DP (DDP) priors (MacEachern, 1999, 2000) can be used for $G_{r,\mathcal{X}}$, $r = 1, 2$

- Briefly, the idea is to use the constructive definition of the DP where now the point masses are i.i.d. realizations from a base stochastic process (say a Gaussian process working with $\log(\sigma_{ri})$), retaining the same (common α) stick-breaking construction for the weights

- A key advantage of the DDP model is its flexibility in capturing different shapes for different covariate values (both observed and unobserved covariate values)

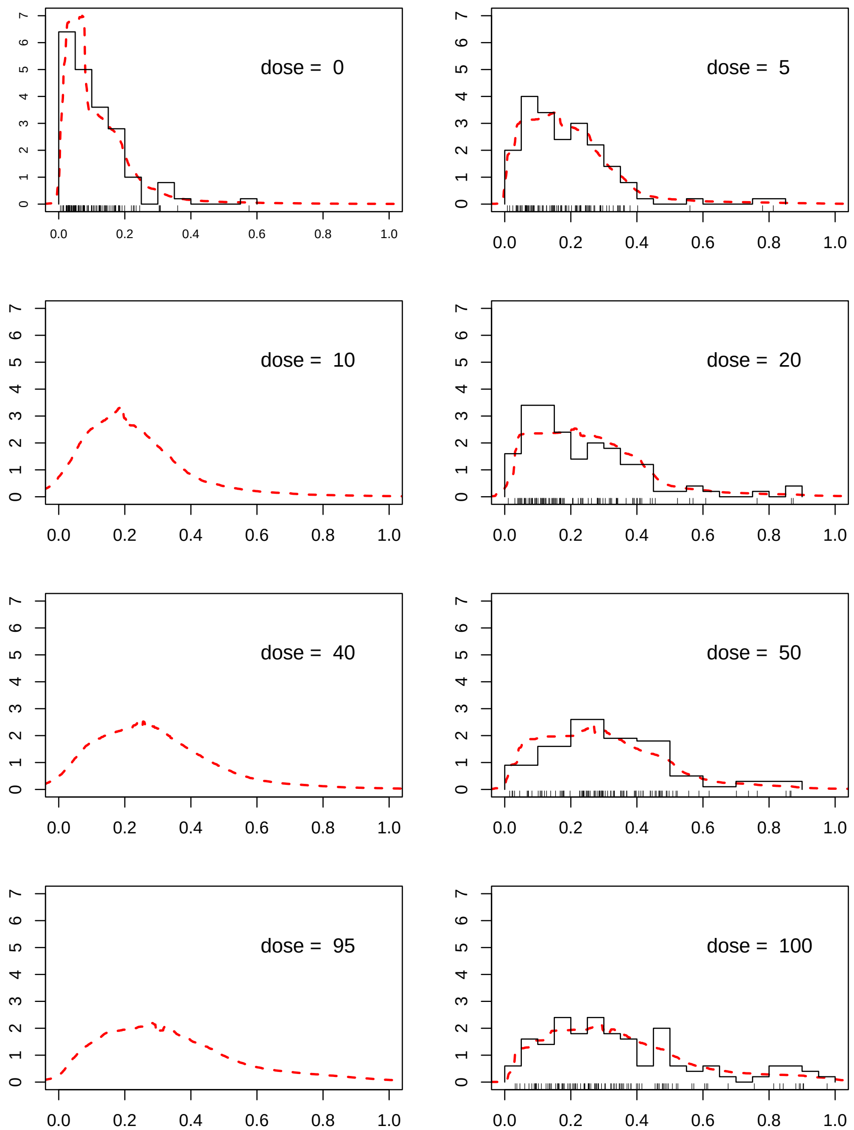
- Several other recent constructions and extensions of the DDP framework: ANOVA DDP, spatial DP, hierarchical DP, ordered DP, nested DP, local DP ...

Bayesian semiparametric quantile regression



Simulation experiment for the DDP quantile regression model. Posterior predictive densities (dashed lines) at five observed covariate values, overlaid on histograms of the corresponding response observations, and at two new covariate values, $x = 10$ and $x = 95$, overlaid on corresponding true densities (solid lines).

Bayesian semiparametric quantile regression



Comet assay data. Posterior predictive densites under the DDP model (dashed lines) at the five observed dose values, overlaid on histograms of the corresponding responses, and at 3 new doses (10, 40, and 95).

4. Fully nonparametric inference for quantile regression

(joint work with Matt Taddy)

- Semiparametric additive quantile regression framework:
  $\longrightarrow$ enables readily interpretable inference by separating quantile regression function from error distribution
  $\longrightarrow$ proposed Bayesian semiparametric model yields flexible inference for unimodal error densities with parametric quantile regression functions

- A possible extension: add nonparametric prior models for the quantile regression functions h_m in the additive setting: $y_i = \sum_{m=1}^{M} h_m(x_{mi}) + \varepsilon_i$
  $\longrightarrow$ current work studies the utility and feasibility of Gaussian process priors for the h_m

Fully nonparametric inference for quantile regression

- Alternative model-based nonparametric approach:
  $\rightarrow$ model joint density $f(y, \boldsymbol{x})$ of the response y and the M-variate vector of (continuous) covariates $\boldsymbol{x}$ with a DP mixture of normals:

$$f(y, \boldsymbol{x}) \equiv f(y, \boldsymbol{x}; G) = \int \mathrm{N}_{M+1}(y, \boldsymbol{x}; \boldsymbol{\mu}, \Sigma) \mathrm{d}G(\boldsymbol{\mu}, \Sigma), \;\; G \sim \mathrm{DP}(\alpha, G_0)$$

  with $G_0(\boldsymbol{\mu}, \Sigma) = \mathrm{N}_{M+1}(\boldsymbol{\mu}; \boldsymbol{m}, V) \times \mathrm{IWish}(\Sigma; \nu, S)$

- For any grid of values $(y_0, \boldsymbol{x}_0)$, obtain posterior samples for:
  $\rightarrow$ joint density $f(y_0, \boldsymbol{x}_0; G)$, and marginal density $f(\boldsymbol{x}_0; G)$
  $\rightarrow$ conditional density $f(y_0 \mid \boldsymbol{x}_0; G)$ and conditional cdf $F(y_0 \mid \boldsymbol{x}_0; G)$
  $\rightarrow$ conditional quantile regression $q_p(\boldsymbol{x}_0; G)$, for any $0 < p < 1$

Fully nonparametric inference for quantile regression

- Approach to inference requires more general posterior simulation methods, which include sampling from the posterior of G (more demanding computationally than the semiparametric model)

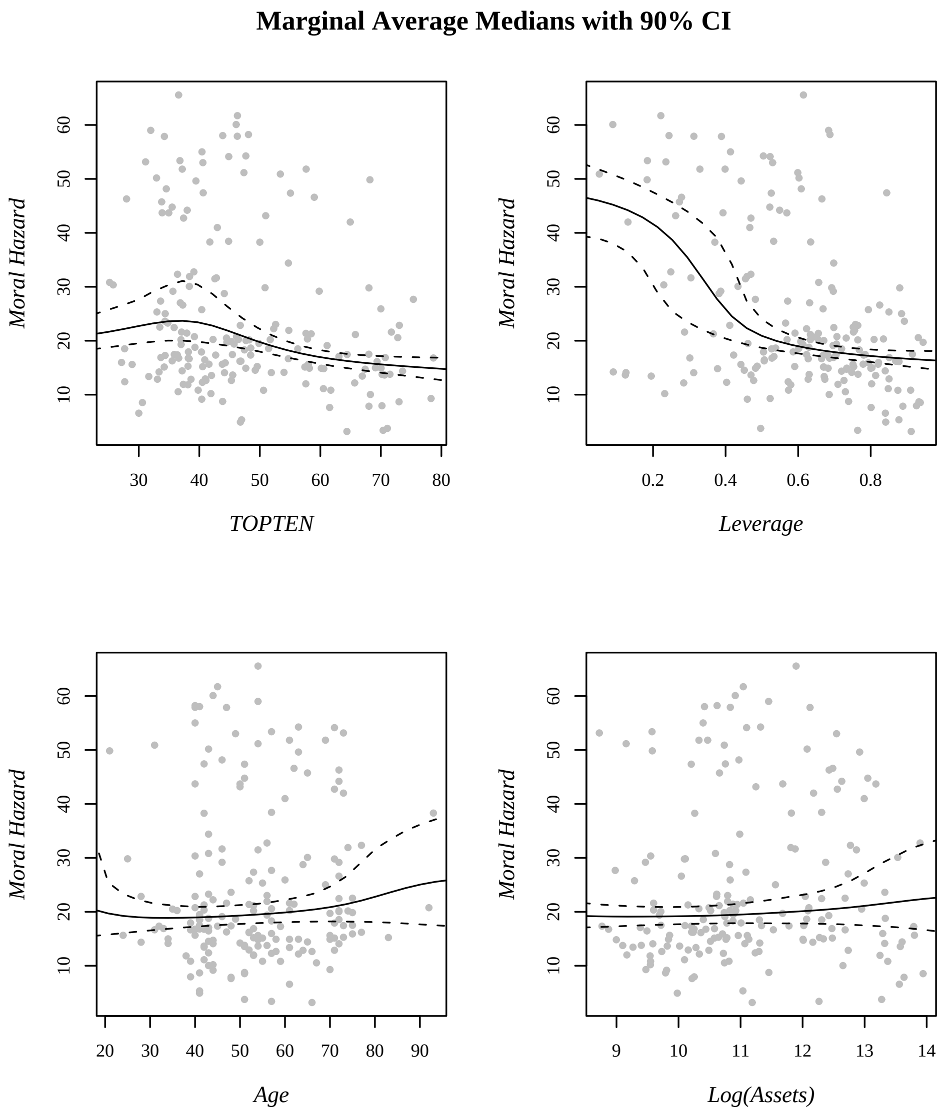
- Key features:
  $\rightarrow$ modeling framework enables simultaneous inference for more than one quantile regression
  $\rightarrow$ model allows flexible response distributions **and** non-linear quantile regression relationships

Data Example

- *Moral hazard* data on the relationship between shareholder concentration and several indices for managerial moral hazard in the form of expenditure with scope for private benefit (Yafeh & Yoshua, 2003)
  $\rightarrow$ data set includes a variety of variables describing 185 Japanese industrial chemical firms listed on the Tokyo stock exchange

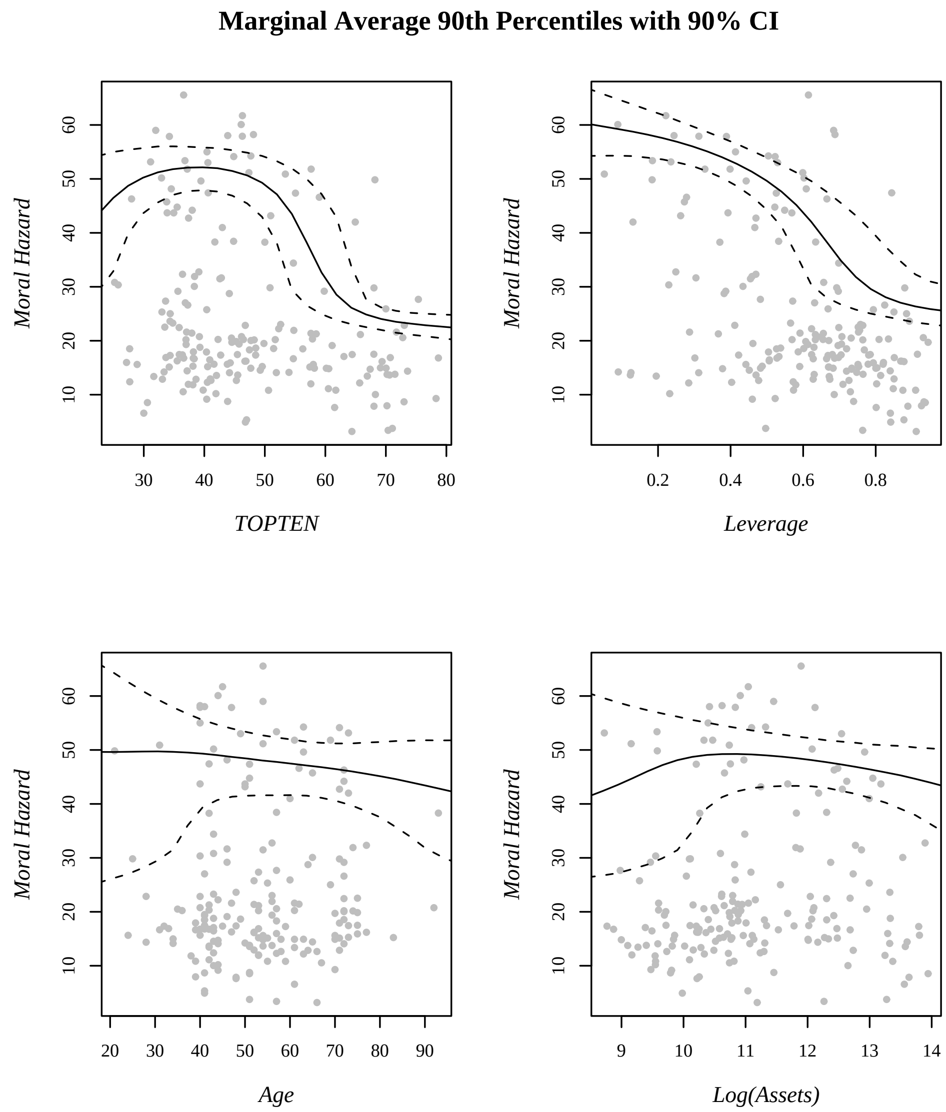
  $\rightarrow$ response y: index *MH5*, consisting of general sales and administrative expenses deflated by sales
  $\rightarrow$ four-dimensional covariate vector $\boldsymbol{x}$: *Leverage* (ratio of debt to total assets); $\log(Assets)$; *Age* of the firm; and *TOPTEN* (the percent of ownership held by the ten largest shareholders)

Fully nonparametric inference for quantile regression



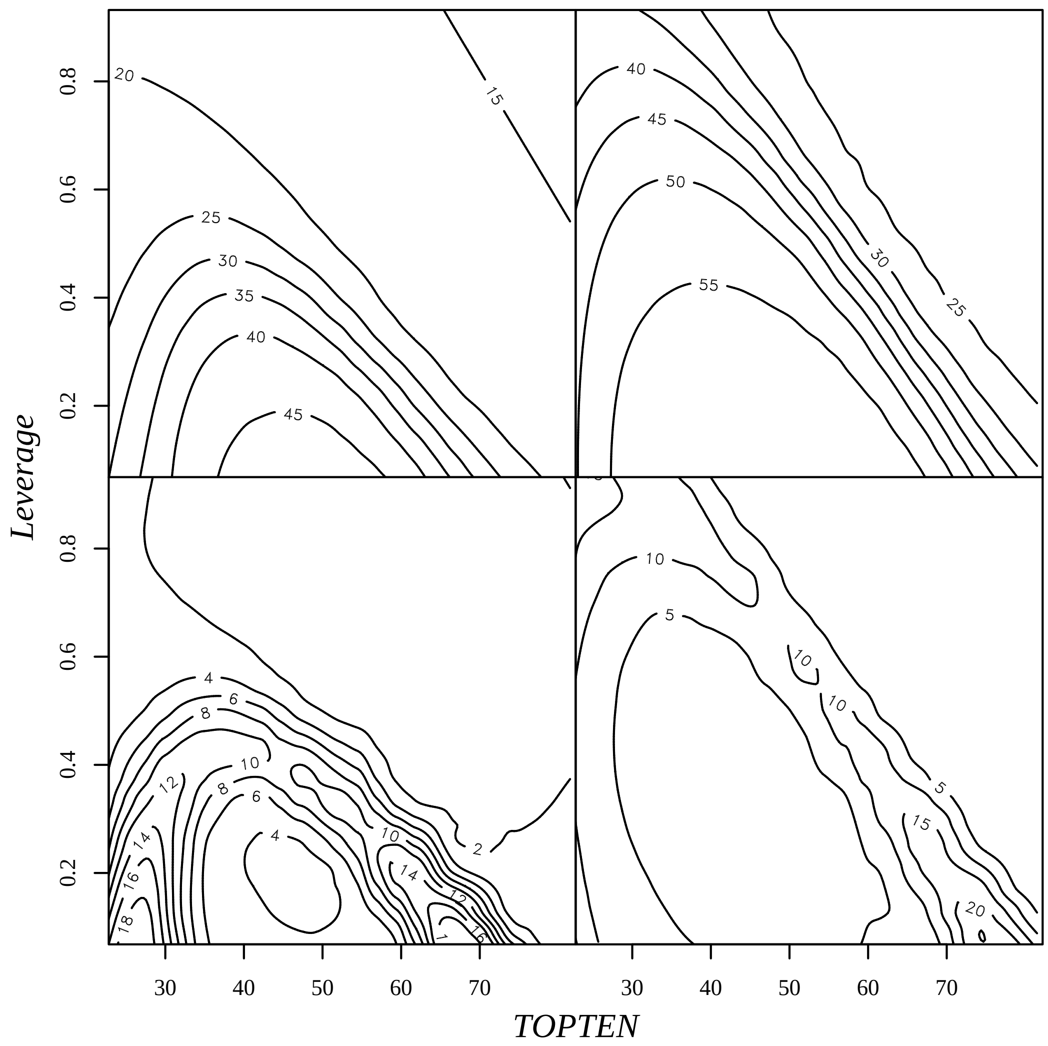
Posterior mean and 90% interval estimates for median regression for $MH5$ conditional on each individual covariate. Data scatterplots are shown in grey.

Fully nonparametric inference for quantile regression



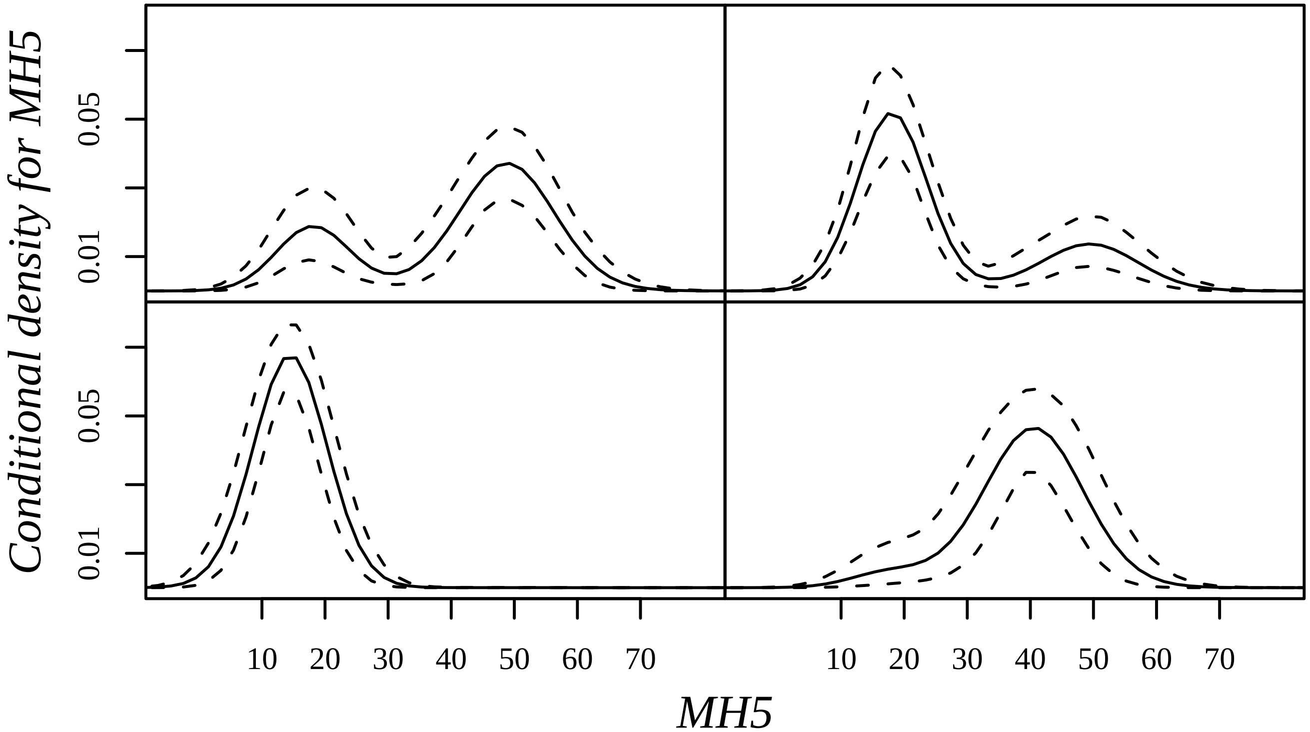
Posterior mean and 90% interval estimates for 90th percentile regression for $MH5$ conditional on each individual covariate. Data scatterplots are shown in grey.

Fully nonparametric inference for quantile regression



Posterior estimates of median surfaces (left column) and 90th percentile surfaces (right column) for $MH5$ conditional on each *Leverage* and *TOPTEN*. The posterior mean is shown on the top row and the posterior interquartile range on the bottom.

Fully nonparametric inference for quantile regression



Posterior mean and 90% interval estimates for response densities $f(y \mid \boldsymbol{x}_0; G)$ conditional on four combinations of values $\boldsymbol{x}_0$ for the covariate vector (*TOPTEN*, *Leverage*, *Age*, $\log(Assets)$) (clockwise from top left, $\boldsymbol{x}_0 = (40, 0.3, 55, 11)$, $(35, 0.6, 55, 11)$, $(40, 0.3, 70, 13)$, and $(70, 0.8, 55, 11)$)

Fully nonparametric inference for quantile regression

- **Model elaborations:**
  → extensions to incorporate both categorical and continuous covariates through mixed discrete-continuous kernels for the DP mixture model
  → modeling for partially observed responses (and/or covariates): quantile regression for survival analysis data with censoring; fully nonparametric Tobit quantile regression for econometrics data

- General framework with potentially important applications beyond quantile regression:
  → nonparametric switching regression modeling
  → modeling and inference for marked spatial Poisson processes
  → sensitivity analysis and inversion of computer model experiments

- Contact info:
  e-mail: thanos@ams.ucsc.edu, web: http://www.ams.ucsc.edu/∼thanos

- UCSC Department of Applied Math and Statistics: www.ams.ucsc.edu
- Technical Reports series: http://www.ams.ucsc.edu/reports/trview.php

Ε Υ Χ Α Ρ Ι Σ Τ Ω !!!